DOCUMENT RESUME

ED 171 789 TM 009 266

AUTHOR	Crane, Laura B.; Cech, Joseph
TITLE	Title I Evaluation Models A1 and B1: An Empirical Comparison.
PUB DATE	Apr 79
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Achievement Gains; Compensatory Education Programs; *Control Groups; Early Childhood Education; *Evaluation Methods; Experimental Groups; Models; National Norms; *Norm Referenced Tests; Predictive Measurement; Program Evaluation; Research Design; Scoring Formulas; *Statistical Bias; Student Testing
IDENTIFIERS	Elementary Secondary Education Act Title I; *Normal Curve Equivalent Scores; *RMC Models

ABSTRACT

Normal curve equivalent achievement gains estimates were compared with RMC Title I evaluation Models A1 and B1. The comparison focused upon the amount of bias introduced by Model A1 when its underlying assumptions were violated. The model assumes, first, that the local school population is accurately represented by the national norm group; and secondly, that the percentile standing of the treatment group on the pretest remained unchanged on the posttest in the abscence of treatment effect. Data were from a Model B1 evaluation--kindergarten children were pre- and posttested with the Aural Comprehension subtest of the Stanford Early School Achievement Test, while the first and second grade children were pre- and posttested with the Woodcock Reading Mastery Test. Comparison of the three model B1 gain estimates (unadjusted, covariance adjusted, and principal-axis adjusted), illustrated the wide variation between estimates. For kindergarten groups, the adjustments were relatively modest, but for grades one and two, the adjustments caused a dramatic shift from negative to positive estimates. Model B1 (adjusted) provided much lower estimates for kindergarten and much higher estimates for grades one and two than did Model A1. Questions were raised concerning representativeness of the norm group and validity of the equipercentile assumption. (Author/CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171789

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Title I Evaluation Models A1 and B1: An Empirical Comparison

Laura R. Crane, Educational Testing Service
Joseph Cech, Schaumburg (Ill.) School District 54

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Laura R. Crane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

paper presented at the annual meeting of
The American Educational Research Association
San Francisco, April, 1979

TM009 266

Introduction

Several local educational agencies have now implemented one or more of the RMC models for Title I evaluation. The three models are referred to as: "Model A, the Norm-Referenced Model; Model B, the Control Group Model; and Model C, the Special Regression Model" (Tallmadge and Wood, 1978, p. 30). The order of preference (in terms of scientific rigor) is given as Model B, Model C, and Model A (Tallmadge and Wood, pp. 22-24). This paper reports the results of a study comparing NCE gain estimated using Model A procedures with NCE gain estimated using Model B procedures. (Technically, the comparison is between Model A1 and Model B1 which requires norm-referenced testing as opposed to Models A2 and B2 which allow for criterion-referenced testing).

The basic difference between models is how the expected no treatment estimate is derived. Model A considers the norm group from a nationally standardized test as a control group and assumes that the percentile standing of the treatment group on a pretest would, in the absence of treatment effect, remain unchanged on a posttest (Tallmadge and Wood). Model B simply uses the posttest percentile standing (converted to NCEs) of the control group for the expected no-treatment estimate (with or without adjustment).

At this point, it is readily apparent, that for Model A1 and Model B1 to yield equivalent NCE gain estimates, two assumptions are critical. The first assumption is that the local school population is accurately represented by the norm group population in terms of relevant characteristics (e.g. SES, ethnic, urban-rural). The second assumption, already

stated, is that the percentile standing of the treatment group on the pretest remain unchanged on the posttest in the absence of treatment effect. Clearly, in a field setting, the first assumption will be violated to some extent and the second assumption has been questioned under any circumstances (Linn, 1978, pp. 8-12). Thus, given a properly implemented control group design, such as Model B, an empirical comparison of Title I NCE gains estimated by applying both Model A1 and Model B1 procedures should provide an indication of the amount of bias that Model A1 procedures can introduce in the estimate of Title I NCE gain.

The key point upon which this comparative study rests is "a properly implemented control group." Ideally, for Model B1, this requires random assignment to treatment and control groups. Since random assignment is frequently not feasible, "random in effect" has been deemed acceptable (Tallmadge and Wood, p. 6). In situations where pretest treatment-control group differences exist, two adjustment procedures (analysis of covariance and principal-axis) are provided (Tallmadge and Horst, 1976). The selection of adjustment procedure depends upon whether or not it can be assumed the two groups are samples from the same population. If the two groups are from the same population, covariance adjustment is appropriate; while if the two groups are from different populations, principal-axis or standardized change score adjustment is appropriate (Tallmadge and Horst, pp. 4-11; Kenney, 1975, pp. 345-361; Linn and Werts, 1977, pp. 229-234.) In any case, Tallmadge and Horst caution that the adjustments are not appropriate if the treatment group and control group regression lines

(or principal-axes) are not parallel such as might occur if the two groups exhibited different growth rates. Keeping these cautions in mind, the following sections: describe the data source and selection of control groups; describe the methods for analysis; present the results and discussion thereof; and present the summary and conclusions of the study.

Data Source

The Schaumburg (Illinois) School District 54 implemented a Model B1 evaluation plan to evaluate Title I programs in kindergarten, first, and second grades. The control groups consisted of children from "comparable" but nonparticipating schools within the district. Identical procedures were followed in selecting Treatment and Control group children and all children were selected by the Director of Research through review of previous testing and/or recommendation of Child Study Teams. From the "pool" of pupils eligible for participation, Title I participants were selected solely on attendance at a school having Title I programs. Control group children were pre- and posttested by substitute teachers trained for the purpose, while treatment group children were tested by classroom teachers.

Kindergarten children were pre- and posttested with the Aural Comprehension subtest of the Stanford Early School Achievement Test (SESAT) while first and second grade children were pre- and posttested with the individually administered Woodcock Reading Mastery Test (WRM). Raw score (RS) data were converted to NCE data through percentile norm tables for the SESAT and first grade WRM tests. For the second grade WRM test, raw scores were converted directly to NCEs using the newly developed 1978 norms.

Because both treatment group and control group children also received assistance from a learning disabilities specialist (LD) and/or a speech therapist (ST) and there was some concern that the children in the control group might have received substantially more assistance than treatment group children, individual pupil data in terms of total minutes spent with each therapist were also collected.

Method for Analysis

To compare Models A1 and B1, NCE gains were estimated according to the procedures recommended for each model. Specifically, for Model A1, gains were estimated according to the formula:

$$NCE_A = NCE_2^T - NCE_1^T \quad (1).$$

For Model B1, both unadjusted and adjusted NCE gains were estimated according to the formulas:

$$NCE_B = NCE_2^T - NCE_2^C, \text{ no adjustment} \quad (2);$$

$$NCE_{Bc} = [NCE_2^T - b(NCE_{(1)}^T - NCE_1^{(T+C)})] - [NCE_2^C - b(NCE_1^C - NCE_1^{(T+C)})]$$

covariance adjustment (3);

$$NCE_{Bp} = NCE_2^T - [NCE_2^C - \frac{SD_2}{SD_1}(NCE_1^C - NCE_1^T)], \text{ principal-axis adjustment (4).}$$

The notation employed in formulas 1-4 is given below:

NCE with an A or B(or B_c, B_p) subscript refer to mean gain estimated according to Model A1 or Model B1 procedures respectively.

T or C superscripts refer to Treatment Group or Control Group means respectively (T + C refers to the overall mean).

1 or 2 subscripts refer to NCE pretest (1) or NCE posttest (2) means.

b is the pooled within groups regression coefficient of NCE_2 on NCE_1.

SD_2 and SD_1 are the pooled within group post- and pretest standard deviations respectively.

Results

NCE means and standard deviations are reported in Table 1. Significant differences ($\alpha < .05$) between groups as determined by exact (or approximate when required) t-tests are also noted in Table 1. There were no significant differences between the kindergarten (Kg) groups on any of the measures. First grade (P1) groups differed significantly on both pretest (RS and NCE) measures. Second grade (P2) groups differed significantly on pre- and posttest RS and NCE measures. Within grade level, none of the treatment-control group comparisons on time spent with speech therapist or learning disabilities teacher were significant (however, the variances were quite disparate in some cases).

Tables 2A, 2B, and 2C report the within group correlations for each of the three grades (Kg, P1, and P2). Also reported are the group-variate correlations. Treatment Group correlations are given below the diagonal and Control Group correlations are given above the diagonal. These results are included to assist in discussion of the NCE gain estimates derived from Model A1 and Model B1 procedures.

NCE gain estimates are given for each of the models in Table 3. Within group slopes for the two adjustment procedures suggested for Model B1 are also provided.

Discussion of Results

The discussion is focused on the Model A1 and Model B1 NCE gain estimates given in Table 3. However, references will be made to the other tables to assist in interpreting the differences between NCE gain estimates.

Comparison of the three Model B1 gain estimates given in Table 3 illustrates the wide variation between estimates depending on whether or not adjusted means were used in computing the NCE gain. For Kg groups, the adjustments were relatively modest but for P1 and P2 groups the adjustments provided a dramatic shift from negative to positive estimates of project impact. The Kg results point to the problems of assessing project impact at that age with these models; not surprisingly the pre-posttest correlations, although significant, are very low, on the order of .3 (see Table 2A). The P1 pre- posttest correlations, are somewhat better on the order of .6 (see Table 2B) while the P2 pre-post-test correlations are quite respectable, on the order of .8 (see Table 2C).

Examination of the within group regression slopes indicated Treatment and Control groups had similar slopes. F-tests for heterogeneous slopes (Winer, 1971, p. 773) revealed no significant differences (Kg, $F_{1,148} < 1$; P1, $F_{1,148} = 1.5$; P2, $F_{1,107} < 1$).

Now changing focus, examine Model A1 NCE gain estimates derived for both Treatment and Control groups. Under the Model A1 equipercentile assumption, Control Group NCE gains should be close to zero which clearly is not the case. Since the differences between Treatment and Control groups on time spent with either speech therapist or learning disabilities teachers were not significant ($\alpha > .05$) it is unlikely that these additional services account for the Control group NCE gain estimates derived under Model A1, presumably, the additional services can be considered part of the regular school program.

Since Model A1 is highly dependent on appropriate testing (e.g. test and level selection, empirical norms, or testing within the empirical norming period) some of the Control Group nonzero Model A1 gains may in fact have been due to testing problems. Both Kg and P1 tests involved interpolated norms, spring for the SESAT (Kg) and fall for the WRM (old norms, P1). Where fall norms are interpolated, a positive bias is likely to be introduced in the NCE gain estimate (Horst, 1978, p. 168); presumably use of spring interpolated norms would result in the reverse, a negative bias. However, the positive Kg Control Group gains (spring interpolated norms) and the negative P1 Control Group gains (fall interpolated norms) would probably not account for the large nonzero gains demonstrated by the Kg and P1 Control Group Model A1 gains. P2 Control Group Model A1 gains were also large and negative (fall and spring norms were empirical). Another possibility at least for P2 (and probably for P1) is that the norms

introduce considerable bias. Strand, Anderson and Sauer (1979, p 15) found that the pre- and posttest empirical norm group raw score means translated to NCEs of 53 and 46.9 respectively resulting in an NCE gain of -6.1 for the P2 norm group. Assuming that the median pre- and posttest raw scores are close to the mean raw scores (or that the distributions are normal), then the P2 WRM norms appear to have a negative bias on the order of -6.1; which when subtracted from the Model A1 Control Group NCE gain results in an NCE gain of -2.6, much more in keeping with an expected zero gain for the Control Group. Finally, Linn cites a study by Kaskowitz and Norwood which suggested that where pretest scores are not particularly low, posttest expected scores will be inflated, (Linn, 1978, p. 11). This also might have influenced the Control Group Model A1 gain estimates since the pretest scores were not particularly low.

Comparing Treatment group NCE gains estimated under Model A1 with those estimated under Model B1 (adjusted) indicates Model B1 provided a substantially lower NCE gain estimate for Kg and substantially higher NCE gain estimates for P1 and P2 than the corresponding Model A1 estimates. An oddity apparent for the three grades is that the difference between Treatment and Control Group's Model A1 estimates is very close to the Model B1 principal-axis or covariance adjusted estimate of Treatment Group gains. This result needs further investigation since it may provide a basis for simplifying Model B1 procedures.

Summary and Conclusions

This study compared Model A1 NCE gain estimates with Model B1 NCE gain estimates. The comparison was intended to indicate the amount

of bias introduced by Model A1 procedures when its underlying assumptions were violated. The data were from a Model B1 Title I evaluation conducted in the Schaumburg (Illinois) District 54 school system for grades Kg, P1 and P2.

Results were equivocal with Kg Model A1 NCE gains considerably larger than Model B1 NCE gains with the reverse ocurring for P1 and P2 NCE gain estimates (adjusted Model B1 NCE gains were greater than Model A1 NCE gains). Some questions were raised concerning representativeness of the norm group and validity of the equipercentile assumption, though not in the usual direction with respect to the equipercentile assumption. The NCE pretest means were somewhat high, especially for the Control Group (P1 and P2 grades) which may have reverse implicatons with respect to the equipercentile assumption.

Test bias was mentioned as a probable problem area. Comparisons of Treatment and Control Group means on Model A1 gain estimates can provide some insight into the size and direction of the bias. Test specific bias can be partially examined through comparison of Norm Group raw score means converted to NCEs. The latter type of bias may "cancel" out in large aggregates across tests, but for within district use of data this becomes an important consideration, especially if the bias is different across levels of a test.

The study is limited in that comparisons for only grades Kg, P1 and P2 were possible. Since the RMC models were intended only for grades P2 and above, this is a rather severe restriction.

A further limitation is that data from only one district were examined. It is quite apparent that results from several studies across grades and tests should be examined before any pattern would be discernable.

Since any Model B1 evaluation data can also be applied to Model A1 procedures if a few precautions are taken (e.g. testing during appropriate times, with respect to norming dates), additional empirical studies could be conducted with relative ease.

Possibly, some simulation studies would be informative by allowing for planned variation in pretest means and ranges of differences between Treatment and Control groups. Certainly parameters affecting bias in Model A1 need further refinement, especially as Model A1 is likely to be the most frequently implemented Model.

TABLE 1

TREATMENT (T) AND CONTROL (C) GROUP RAW SCORE (RS) AND NCE MEANS AND STANDARD DEVIATIONS (SD) FOR KINDERGARTEN (Kg) FIRST (P1) AND SECOND (P2) GRADE PUPUILS

	Group	N	Pre RS Mean	(SD)	Post RS Mean	(SD)	Pre NCE Mean	(SD)	Post NCE Mean	(SD)	Time (Min.) With Speech Therapist Mean	(SD)	Learning Disabilities Teacher Mean	(SD)
Kg	T	82	13.5	(3.49)	19.1	(3.37)	39.0	(15.41)	47.0	(14.82)	106.9	(226.16)	41.4	(86.60)
	C	70	14.6	(3.63)	19.6	(4.11)	43.9	(16.28)	50.0	(18.98)	146.6	(294.76)	50.9	(176.15)
P1	T	83	39.2	(5.42)[a]	69.2	(14.37)	48.2	(10.20)[a]	36.3	(20.78)	116.0	(259.66)	113.5	(287.63)
	C	69	43.3	(7.08)	70.4	(12.51)	55.9	(12.87)	37.8	(18.13)	166.0	(246.12)	139.0	(345.90)
P2	T	46	77.9	(38.51)[a]	149.7	(50.00)[a]	34.2	(19.58)[a]	35.8	(19.76)[a]	42.6	(143.46)	177.4	(648.93)
	C	65	127.3	(46.01)	175.1	(39.79)	54.1	(15.85)	45.4	(16.13)	71.6	(171.45)	101.1	(264.61)

[a] $\alpha < .01$

TABLE 2A

KINDERGARTEN PRE-POSTTEST CORRELATIONS (RAW SCORE AND NCE) AND CORRELATIONS FOR TIME SPENT WITH A SPEECH THERAPIST (ST) OR LEARNING DISABILITIES TEACHER (LD)[a]

	Group[b]	PreRS	PostRS	PreNCE	PostNCE	Time w/ Speech Therapist	Time w/ Learning Disabilities Teacher
PreRS	.150		.373[d]	.992[d]	.345[d]	.017	.040
PostRS	.068	.322[d]		.355[d]	.992[d]	.162	.109
PreNCE	.152	.996[d]	.331[d]		.327[d]	.041	.043
PostNCE	.090	.323[d]	.997[d]	.332[c]		.154	.087
ST	.076	-.060	-.183[c]	-.052	-.176		-.016
LD	.035	.049	-.004	.076	.001	.225[c]	

[a]The correlations within the square section as marked are within group correlations; Treatment Group correlations are given in the lower half of the matrix while Control Group correlations are given in the upper half of the matrix.

[b]The correlations under the column heading "Group" are correlations between group membership (treatment or control) and the variables given as row headings.

[c]$\alpha < .05$

[d]$\alpha < .01$

TABLE 2B

FIRST GRADE PRE-POSTTEST CORRELATIONS (RAW SCORE AND NCE) AND CORRELATIONS FOR TIME SPENT WITH A SPEECH THERAPIST (ST) OR LEARNING DISABILITIES TEACHER (LD) [a]

	Group[b]	PreRS	PostRS	PreNCE	PostNCE	Time w/ Speech Therapist	Time w/ Learning Disabilities Teacher
PreRS	.313[d]		.626[d]	.998[d]	.627[d]	-.060	-.041
PostRS	.045	.575[d]		.631[d]	1.000[d]	-.182	-.207[c]
PreNCE	.319[d]	.998[d]	.578[d]		.631[d]	-.070	-.040
PostNCE	.040	.572[d]	.999[d]	.575[d]		-.185	-.207[c]
ST	.098	.095	.079	.091	.078		.066
LD	.040	-.176	-.223[c]	-.180	-.227[c]	.268[d]	

[a] The correlations within the square section as marked are within group correlations; Treatment Group correlations are given in the lower half of the matrix while Control Group correlations are given in the upper half of the matrix.

[b] The correlations under the column heading "Group" are correlations between group membership (treatment or control) and the variables given as row headings.

[c] $\alpha < .05$

[d] $\alpha < .01$

TABLE 2C

SECOND GRADE PRE-POSTTEST CORRELATIONS (RAW SCORE AND NCE) AND CORRELATIONS FOR TIME SPENT WITH A SPEECH THERAPIST (ST) OR LEARNING DISABILITIES TEACHER (LD)[a]

	Group[b]	PreRS	PostRS	PreNCE	PostNCE	Time w/ Speech Therapist	Time w/ Learning Disabilities Teacher
PreRS	.495[d]		.886[d]	.982[d]	.883[d]	-.223[c]	-.373
PostRS	.254[d]	.825[d]		.883[d]	.997[d]	-.150	-.363[d]
PreNCE	.492[d]	.950[d]	.824[d]		.885[d]	-.273[c]	-.404[d]
PostNCE	.260[d]	.832[d]	.992[d]	.826[d]		-.149	-.386[d]
ST	.089	.336[c]	.123	.208	.155		.057
LD	-.085	-.076	-.272[c]	-.136	-.193	.025	

[a] The correlations within the square section as marked are within group correlations; Treatment Group correlations are given in the lower half of the matrix while Control Group correlations are given in the upper half of the matrix.

[b] The correlations under the column heading "Group" are correlations between group membership (treatment or control) and the variables given as row headings.

[c] $\alpha < .05$

[d] $\alpha < .01$

TABLE 3

MODEL A1 AND MODEL B1 NCE GAIN ESTIMATES FOR KINDERGARTEN (Kg), FIRST (P1) AND SECOND (P2) GRADES

Grade	NCE Gain Model A1[a] (Formula 1)	NCE Gain Model B1 Unadjusted[b] (Formula 2)	NCE Gain Model B1 Covariance[c] Adjusted (Formula 3)	NCE Gain Model B1 Principal-axis[d] Adjusted (Formula 4)
Kg	8.0 (6.1)	-3.0 (4.9)	-1.3 (.32, .38)	2.2 (.96, 1.17)
P1	-11.9 (-18.1)	-1.5 (7.7)	6.2 (1.17, .89)	12.0 (2.04, 1.41)
P2	1.6 (-8.7)	-9.6 (19.9)	7.6 (.82, .90)	10.8 (1.01, 1.02)

[a] Given in parentheses are the Control Group gains determined according to Model A1 procedures.

[b] Given in parentheses are the pretest NCE differences between the Control and Treatment groups ($NCE_1^c - NCE_1^t$).

[c] Given in parentheses are the within group regression slopes for Treatment and Control Groups respectively.

[d] Given in parentheses are the within group principal-axis slopes for the Treatment and Control Groups respectively.

REFERENCES

Kenny, D.A. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin 1975 82 (3) 345-362.

Horst, D.P. Checklists of potential errors in the ESEA Title I evaluation and reporting system. RMC, 1978 UR 331 151-181.

Linn, R.L. The validity of inferences based on the proposed Title I evaluation models. Paper presented at annual meeting, American Educational Research Association, Toronto, Canada, 1978.

Linn, R.L. & Werts, C.E. Analysis implications of the choice of a structural model in the nonequivalent control group design. Psychological Bulletin, 1977 84 (2) 229-234.

Strand, T., Anderson, C., & Sauer, E.J. Exploratory study of the use of a wide range achievement test with Model A in Title I evaluation. Paper presented at the annual meeting of The American Educational Research Association, San Francisco, April, 1979.

Tallmadge, G.K. & Horst, D P. A procedural guide for validating achievement gains in educational projects. Washington, D.C. U.S. Government Printing Office, 1976. (Stock No. 017-080-01516)

Tallmadge, G.K. & Wood, C.T. Users guide: ESEA Title I evaluation and reporting system (revised draft). Washington, D.C.: U.S. Office of Education, 1978.

Winer, B.J. Statistical principles in experimental design. McGraw-Hill, Inc. 1971.